



Mixed-resolution HEVC based multiview video codec for low bitrate transmission

Bruhanth Mallik¹ · Akbar Sheikh-Akbari¹ ·
Ah-Lian Kor¹

Received: 21 September 2017 / Revised: 17 April 2018 / Accepted: 15 June 2018
© The Author(s) 2018

Abstract There has been increasing demand for multiview video transmission over band limited channel over past years and various techniques have been proposed to fulfil this need. In this paper, a High Efficiency Video Codec (HEVC) based spatial resolution scaling type of mixed resolution coding model, MRHEVC-MVC, for frame interleaved multiview videos is presented. However, enabling the HEVC to encode video with different frame resolutions is a challenge due to the coding tree partitioning used by the codec. This has been overcome by super-imposing the low resolution replica of each full resolution frame on their respective decoded picture buffer and setting the remaining space of the frame buffer to zero. The codec's reference frames structure is designed to efficiently encode frame interleaved multiview videos using a HEVC based mixed resolution codec. The proposed MRHEVC-MVC codec has been tested against the standard multiview extension of high efficiency video codec (MV-HEVC) for “Balloon”, “Newspaper1”, “Undo_Dancer”, “Kendo” and “Poznan_Street” standard multiview video sequences. Results show that the proposed codec gives significantly higher coding performance to that of the MV-HEVC codec at low bitrate both subjectively and objectively.

Keywords Texture3D videos · Multiview video coding · Mixed-resolution video codec · HEVC · MV-HEVC · Low bitrate transmission

1 Introduction

Advances in video technologies have played a key role in facilitating applications (such as video conferencing, e-learning, video-on-demand, real time surveillance, automation, robotics, and machine vision) in our day to day activities. The growing influence of these entertainment

✉ Bruhanth Mallik
b.mallik6347@student.leedsbeckett.ac.uk

¹ School of Computing, Creative Technology & Engineering, Leeds Beckett University, Headingley Campus, Leeds LS6 3QS, UK

and interactive applications, ranging from low resolution to ultra-high definition (UHD) video contents, has resulted in steady increase in fixed and mobile network's traffic. According to Cisco's Visual Networking Index (VNI) forecast, it is anticipated that: by 2020, 71% of the overall IP traffic used by both business and domestic users would be from non-PC devices, of which 82% would be for visual content communication [30]. Lately, with the growing popularity of video applications, such as video-on-demand (VoD) and introduction of UHD internet protocol televisoin (IPTV) video streaming, it is expected that these applications will constitute 21% of global VoD traffic by 2020 [ibid]. Consequently, while the quality of network connectivity and bandwidth are improving, demands for high definition (HD) and UHD quality monoscopic and three dimensional (3D) visual communications continue to rise at the same time. Over the last decade, numerous 3D video technologies, such as holography, augmented and virtual reality, stereoscopic, video with depth information coding and multiview video coding, have been studied (e.g. [7, 14, 26]). Of the various 3D video technologies, multiview video's capability to provide multiple viewing angles or points of dynamic scenes with high quality and immersive multimedia experience, makes it a key technology for many texture-related 3D applications (e.g. free-viewpoint video, 3D-TV, and immersive teleconferencing). The multiview version of the 3D videos uses multiple cameras to capture the same scene concurrently by geometrically aligned and synchronized cameras [26]. Due to huge amount of visual information in multiview videos, the bitrate needed to encode these videos increases approximately linearly with the number of views used to reconstruct the 3D content. To efficiently compress these huge amount of visual information, multiview video codecs (MVC) depend on extensively exploiting inter-view correlations (in addition to spatial and temporal correlations within each view) within multiview video sequences [7]. MVC largely benefits from disparity prediction/ compensation to exploit correlations between frames of the neighbouring views and thus provides higher compression efficiency compared to coding individual views independently. However, the amount of buffer memory that is made feasible by MVC, for additional decoded pictures, restricts the number of views that can be encoded concurrently [12, 14]. Hence, to efficiently implement MVC, these challenges and requirements of the codec have been extensively studied and standard texture based multiview point-to-point codecs such as: H.264/MPEG-4 MVC [12, 29] and multiview extension of high efficiency video codec (MV-HEVC) [28] have been developed. MV-HEVC's design is based on MVC extension of H.264/ MPEG-4 AVC framework. The high level syntax of the MV-HEVC codec is an extension of the H.264/ MPEG-4 AVC codec. Although MV-HEVC's toolsets are designed to deliver reduced signaling overhead for motion information, its layer architecture for representing dependent views of multiview video stagnates the compression efficiency [28]. On the other hand, achieving high compression ratios using large quantization parameters leads to distortions and blocking artefacts, due to loss of high frequency information within the videos. Hence, multiview video transmission over a bandwidth limited communication channel demands high level of compression efficiency.

To overcome the setbacks encountered by multiview video codecs at low bitrate transmission, mixed-resolution based multiview video codecs, also known as asymmetric spatial inter/intra view frames resolution video codecs, are proposed in [10, 16]. A subjective study of the quality of spatially and temporally down-sampled stereo video frames, shows that spatially down-sampled videos are preferred and are potentially more suitable for stereoscopic video coding at low bitrate [21]. Further studies have shown that mixed-resolution based 3D video coding in case of stereoscopic videos for low bitrate transmission offers significant bandwidth savings while retaining adequate video quality [2, 27]. Improving compression efficiency of mixed-

resolution 3D video codecs, using inter-view prediction between frames of different resolutions, is attempted in [6, 9]. A study conducted in [32], to analyse the factors influencing correlations in multiview video sequences, has shown that a single prediction structure cannot be universally effective at all time and for all scenes. Therefore, an adaptive mode switching algorithm called AMVC-HBP based on the standard MVC extension of H.264/AVC is proposed in [32]. To adaptively change the prediction framework, AMVC-HBP depends on the Lagrangian cost of the coded pictures that represent the correlation characteristics of the multiview videos [31, 32]. Despite the incurred long delays and high memory consumption of AMVC-HBP, the technique delivers very little coding gain, as the adaptive prediction framework does not fully utilise inter-view and temporal correlations in less dense and slow moving sequences. A mixed-resolution multiview video coding technique, which uses the statistics of block matching to adaptively reorder the reference frames for multiview extension of the H.264/AVC, is proposed in [19]. This mixed-resolution multiview video codec has been improved by integrating the prediction architecture technique, which is developed in [17] using an adaptive reference frame reordering algorithm reported in [18]. This approach to mixed-resolution multiview video coding has demonstrated some bitrate saving, however, this coding technique uses just one low resolution view in a three view multiview scenario. Since this technique uses an adaptive reference frame re-ordering algorithm that focuses on predicting hard scene changes, the inter-view prediction is limited. Further, this codec's prediction architecture does not demonstrate the potential to fully exploit the inter-view correlations, hence results in lesser coding gain. An inter-view motion vector prediction technique for MV-HEVC using temporal motion vectors from previously encoded motion information is proposed in [20]. This method calculates a global disparity value by referring to look-up table of disparity conversion, which is calculated from previously encoded frames and uses this value to modify the motion field of inter-view reference pictures. An adaptive spatial mixed resolution technique using MV-HEVC is employed in [1] for stereo video coding. This technique uses the correlation between frequency power spectrums of the adjacent views and quality metrics of the encoded down-sampled video frames to choose the appropriate down sampling factor. In [8], the application of HEVC for mixed resolution stereoscopic video coding, aimed at reducing quality-imbalance problem of the decoded video sequences to less than 2 dB, is reported. The prediction structure of this codec shows that the left view video frames are coded without reference to the right view frames. The right view video frames use the adjacent neighboring view frames in addition to the similar left view prediction frames for prediction. In addition, the first frame of the right view in each group of pictures is coded in full resolution. This codec reduces the quality-imbalance between the frames of the neighboring views to less than 2 dB. However, by not fully exploiting of inter-view frame correlation and coding the first frame of the right view in each Group of Pictures (GOP) with full resolution, the capability of mix-resolution stereo video coding with HEVC is underutilized.

In this paper, a spatial resolution scaling based mixed resolution coding architecture for multiview videos using the standard HEVC video codec is proposed. The proposed coding model applies mixed spatial resolution coding to multiview video frames that are reordered using a frame inter-leaving algorithm developed for multiview videos in [13]. The HEVC standard video codec has been modified and configured to encode the frame interleaved multiview videos frames with mixed spatial resolution. Experimental results, for coding five standard multiview test video sequences, show that the proposed video codec generates significantly higher coding performance than the anchor MV-HEVC codec. The rest of this paper is organized as follows: Section 2 presents the framework of the proposed technique by introducing the structure of the mixed resolution multiview videos, the proposed intermediate frame down-sampling method,

multiview video frame interleaving algorithm and the design of the codec. Section 3 includes the experimental results and finally, the paper is concluded in Section 4.

2 MRHEVC-MVC proposed technique

The proposed mixed resolution HEVC based multiview video codec (MRHEVC-MVC) is developed within the standard HEVC codec's software framework. The frame structure of the proposed codec is shown in Fig. 1. The MRHEVC-MVC's design allows key frames (encoded as I-frames) to retain their full resolution, whereas intermediate frames (encoded as P- or B-frames) are spatially down-sampled by a factor of 2. From the proposed coding frame structure, it can be noted that the key frames are mostly selected from the center view in order to maintain a balanced visual quality between the corresponding left and right view frames. A Blackman 2D FIR low-pass filter (filter coefficients are shown in Table 1) is used to mitigate aliasing artefacts due to down-sampling [24].

The filtered intermediate frames are both horizontally and vertically down-sampled to maintain the same aspect ratio as the full resolution key-frames (I-frames). The filtered and down-sampled low resolution intermediate frames are superimposed on top left quadrant of a full resolution sized frame with every pixel value (in other quadrants) set to zero, as illustrated in Fig. 2.

As a result, the intermediate and the key frames of the mixed-resolution video sequence have the same frame size. These pre-processed intermediate frames contain scene information at a lower resolution compared to that of the key frames and are placed in the top-left quadrant of the frame. Since HEVC codec exploits the spatial correlations when encoding I-frames, the regions with zero value pixels are encoded with minimum signaling bits. The mixed resolution multiview video frames are reordered into a monocular video sequence using frame interleaving algorithm shown in Fig. 3 [13].

The frame interleaving algorithm reorders the multiview video frames, such that at least two frames of each view are always adjacent to each other, thereby maximizing exploitation of the temporal correlations, which in turn improves the coding performance of the codec. Frame interleaving mixed resolution multiview video frames of the proposed codec facilitates the exploitation of temporal and interview correlations within the multiview video frames. As the multiview video frames are temporally interleaved, the proposed codec's reference frame structure allows cross-frame (also called as lateral frame) referencing [14], as shown in Fig. 4. It is worth to mention that the standard MV-HEVC does not support cross-frame referencing.

The proposed codec is developed on a standard HEVC codec's platform, where the platform has been modified and configured to encode frame interleaved mixed-resolution

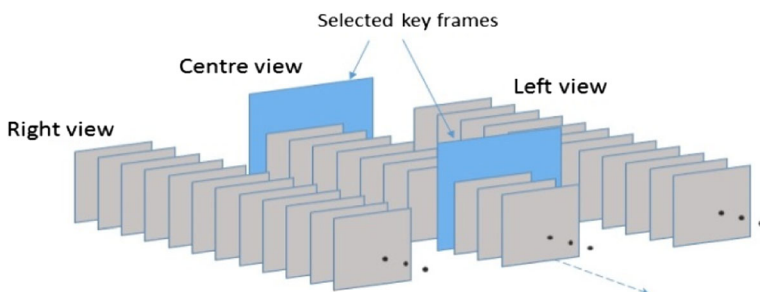


Fig. 1 Frame structure of the proposed mixed resolution multiview video codec

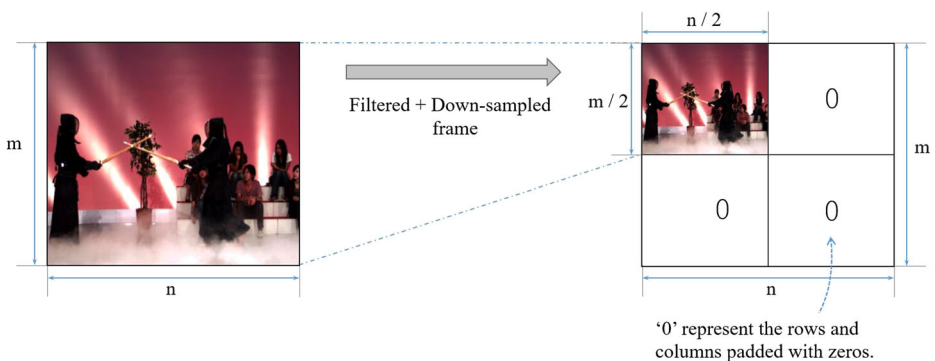
Table 1 Blackman 2D FIR filter coefficients

0.0381	0.1051	0.0381
0.1051	0.4273	0.1051
0.0381	0.1051	0.0381

multiview video frames. The proposed MRHEVC-MVC codec is also designed to be able to use frames of different resolutions for motion estimation/disparity compensation during inter- and intra-view frame referencing. The decoded low resolution video frames are enlarged using Bi-cubic interpolation method to retrieve the original video frame size, as shown in Fig. 5.

One of the primary challenges in implementing a mixed resolution video codec is that the codec should be able to encode frames with different spatial resolutions in the same sequence. Unlike the previous standard codecs of ITU-T and ISO/IEC (such as, H.263, H.264, H.264/AVC) that use macroblocks as the basic processing unit, the standard HEVC codec uses square blocks of Coding Tree Units (CTU). Each CTU represents one luma and two chroma Coding Tree Blocks (CTB), along with their associated syntax. CTUs are further partitioned into the Coding Units (CU) and this subdivision is specified in the CTU through their respective quadtree syntax. The CUs in a CTU are coded in a z-scan order, as shown in Fig. 6 [11, 28].

Similar to CTU, a CU also contains luma and chroma details and its associated syntax. The HEVC splits the input video frames into sequence of CTUs, called Slices. CTUs in a slice are processed in a raster scan. Each slice header contains all coding details of the slice, such as type of the picture coding (either of type I, P or B), identifiers and address of the slice segment, least significant bit of the Picture Order Count (POC), reference picture index, coding tool parameters, initial slice quantization parameter (QP) value. The HEVC's reference picture management (called Reference Picture Set (RPS) in ITU-T and ISO/IEC documentation [11]), uses POC indexing to identify individual reference frames in the Decoded Picture Buffer (DPB). The HEVC's DPB has two segments for reference frames, called L0 and L1, to facilitate bi-directional prediction. The current and a pre-defined number of previous decoded pictures are stored in the DPB. The proposed MRHEVC-MVC codec's reference frame structure, is designed by taking into consideration that the mixed resolution multiview video frames have been reordered into a monocular sequence, starting from the middle view. The reference frame structure of the proposed multiview video codec for coding 3-view scenario is shown in Fig. 4. A mixed resolution based video codec is designed to be able to use video frames with different resolutions as reference frames. However, this poses another problem, i.e. how the inter-frame referencing can be achieved while accurately finding the best match for

**Fig. 2** Intermediate frame resolution downscaling

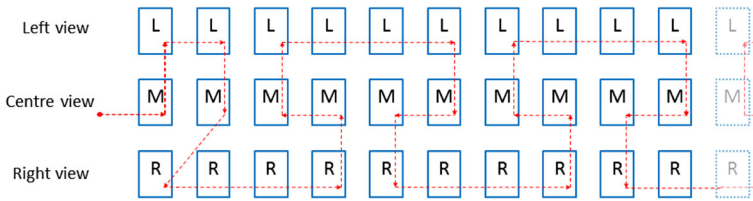


Fig. 3 Frame interleaving algorithm's path to reorder multiview video frames [13]

motion information when the information in the frames are at different resolutions. To address this issue, the proposed codec saves a copy of the decoded I-frame locally, in its full resolution, along with its status flag and CTU details (from the slice header) outside of the Decoded Picture Buffer (DPB). In HEVC, the motion prediction block, intra-frame prediction and in-loop deblocking filtering are all CTB based. The HEVC reconstruction function, which is also available to the deblocking filter, is used in conjunction with DPB in the proposed codec to generate a low resolution replica of the I-frame, as shown in Fig. 2. The coding tree algorithm splits the resulting low resolution I-frame into CTUs and updates the I-Slice header, accordingly. Then the low resolution replica of the I-frame is placed in the memory buffer of the DPB, instead of its full resolution I-frame. The information in the low resolution CTUs of the I-frame in DPB has the same resolution as that of the intermediate frames. This enables the codec to find the motion vectors during inter-frame referencing. The proposed codec constantly checks the status of the picture output flag of the decoded I-frame. If the I-frame's CTUs are no longer needed, the full resolution copy of the I-frame that has been saved outside the DPB is output to the DPB. A flow chart of the algorithm to generate low resolution replicas of I-frames is shown in Fig. 7.

The proposed MRHEVC-MVC codec's inter-frame referencing is designed to allow low-resolution intermediate frames to use other low resolution intermediate frames and also down-sampled replicas of the I-frames. An example of the frame referencing architecture of the proposed codec is illustrated in Fig. 8.

Thus, by making the above mentioned design modifications to the standard HEVC codec, the modified codec is equipped to encode mixed-resolution multiview video frames. The configuration parameters of the HEVC codec are set to enable the implementation of the MRHEVC-MVC codec's reference frame structure, as in Fig. 4.

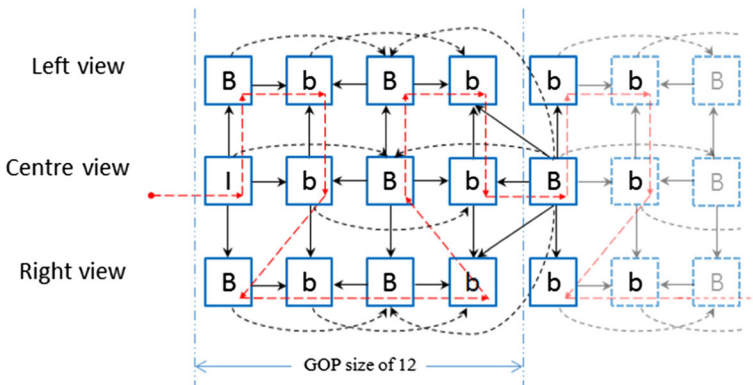


Fig. 4 Reference frame structure of the proposed MRHEVC-MVC codec

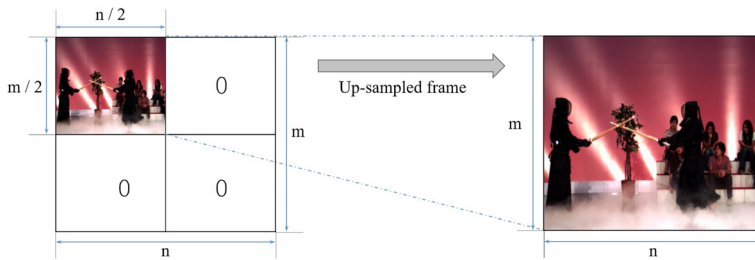


Fig. 5 Intermediate frame resolution up-sampling

To implement the proposed MRHEVC-MVC video codec, JCT-VC's HEVC software version HM16.12 has been used. In addition to modifications to the standard HEVC, the configuration parameters of the codec are set to encode frame interleaved mixed-resolution multiview video frames with the reference frame structure, as shown in Fig. 4. Since the video frames of the three views are reordered using the proposed frame interleaving algorithm, the Group of Pictures (GOP) size is set to 12 frames. Experimental results presented in this paper are generated when intra frame iteration for the proposed codec is set to after 48 frames. The temporal and inter-view correlations of the multiview video sequences are influenced by various parameters, e.g. video content, scene illumination changes, relative velocity of the objects in the scene, camera movement, inter-camera angles and frame rates. The goal of motion estimation/disparity compensation is to reduce the entropy of the difference block, hence the number of bits required to code the coefficients in the block has to lower. This is achieved by finding the best match of the coding unit in either one of the neighboring or previous frames. In case of neighboring view, the location of the best match is a function of the distance of the camera from the scene and inter-camera angles [5, 25]. For the standard test videos used in this study, the search area to find the best match is set to 96 pixels to mitigate the effect of the inter-camera angles and camera distance from the scene for both disparity and motion estimation/compensation.

3 Experimental results

The compression efficiency of the proposed codec is compared with the standard MV-HEVC codec. To evaluate the coding performance of the proposed mixed resolution HEVC based

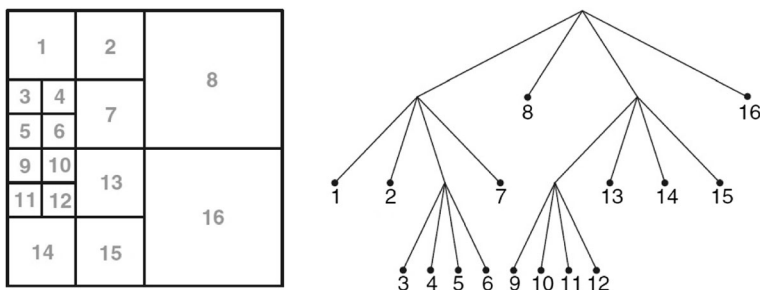


Fig. 6 An example of HEVC's 64X64 coding tree unit (CTU) partitioning into coding units (CUs) of 8 × 8 to 32X32 luma samples [11, 28]

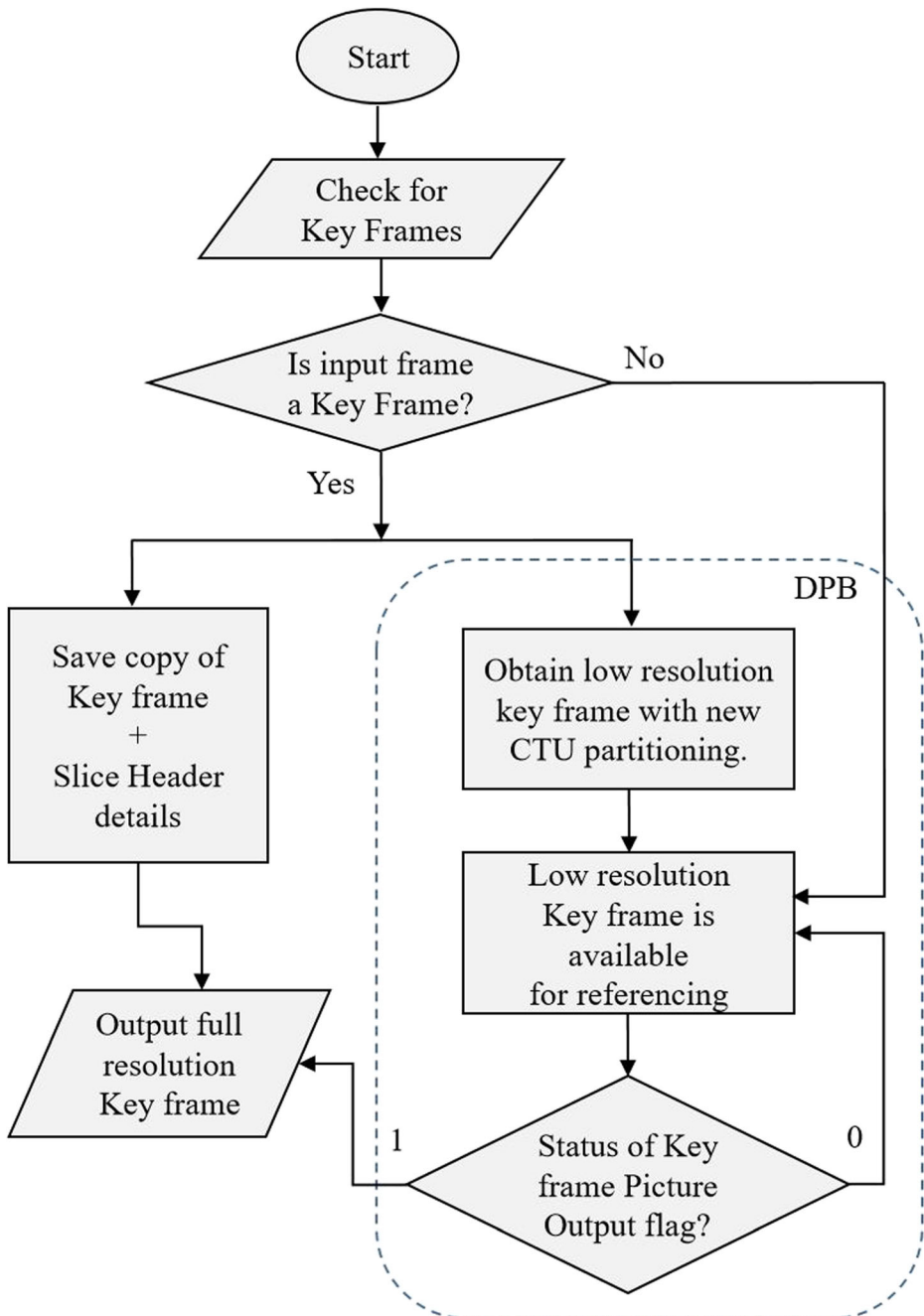


Fig. 7 Flow chart of Key frame down-sampling function

multiview video codec, three views of the standard multiview video datasets of 4:2:0 format, namely “Balloon”, “Newspaper1”, “Undo_Dancer”, “Kendo” and “Poznan_Street” sequences for the views 1-3-5, 2-4-6, 1-5-9, 1-3-5 and 5-4-3 respectively, are selected and coded

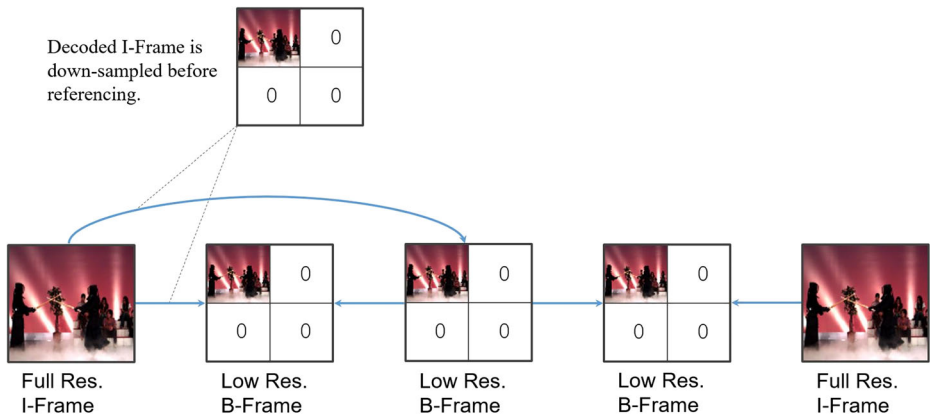


Fig. 8 Example of inter-frame referencing model of the proposed mixed resolution multiview video codec

using the proposed codec. The standard datasets have been produced by recording under different lighting conditions, camera distances and with varying scene characteristics, similar to entertainment and interactive applications. These five standard video datasets cover scenes with both static and dynamic backgrounds with varying level of illumination (Fig. 9 shows the first frame of these five datasets).

The experimental results are compared with the anchor MV-HEVC codec for the five multiview datasets (3 view scenario) that are available in the JCT3V-G1100 common test condition (CTC) documentation [15]. The proposed codec enlarges the encoded low resolution video frames to their original resolution using the Bi-cubic interpolation technique, as shown in Fig. 7. The Peak Signal to Noise Ratio (PSNR) measure is then used to assess the quality of the decoded frames. The combined PSNR (YUV-PSNR) weighted sum of the average PSNR per video frames of the individual components (PSNR-Y, PSNR-U, and PSNR-V) of the

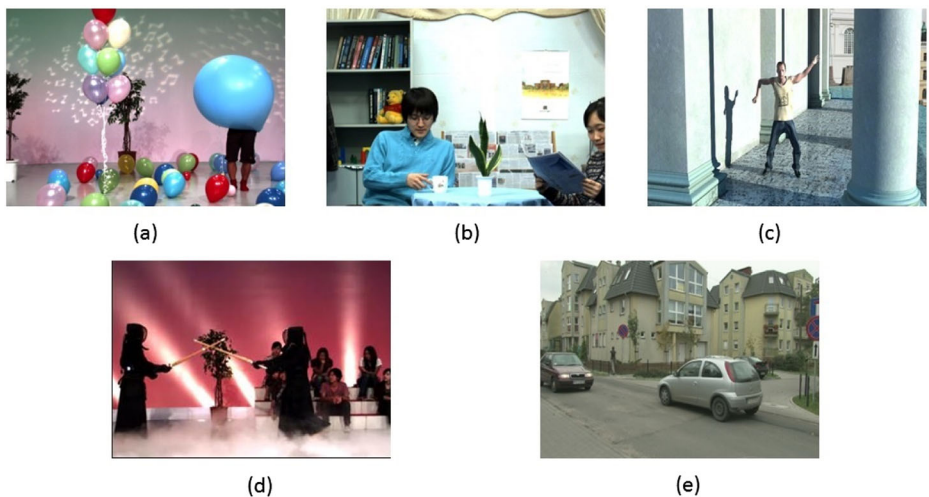


Fig. 9 First frame of the multiview datasets: **a** “Balloon”, **b** “Newspaper1”, **c** “Undo_Dancer”, **d** “Kendo” and **e** “Poznan_Street”

decoded and enlarged multiview videos, as defined in Eq. (1) [22], were calculated and compared with that of the anchor MV-HEVC codec.

$$YUV\ PSNR = \frac{(6\ Y\ PSNR + U\ PSNR + V\ PSNR)}{8} \quad (1)$$

Figures 10, 11, 12, 13 and 14 show the resulting YUV-PSNRs at Quantization Parameters (QP) of 20, 25, 30, 35, 40 and 25, 30, 35, 40 for the proposed HEVC-MRSVC and the anchor MV-HEVC codecs, respectively.

Figure 10, shows the resulting average YUV-PSNR for the “Balloons” multiview videos. The “Balloons” dataset is recorded under multiple artificial lighting conditions, where the background is progressively changing with a number of fast moving objects in the foreground. From Fig. 10, it can be seen that the proposed MRHEVC-MVC codec gives superior coding performance to that of the anchor codec, when transmitting at lower bitrates (1.6 to 0.42dBs higher PSNR than the anchor codec from 265.46 kbps to 500 kbps). Moreover, the proposed codec, generates acceptable video qualities at bitrates lower than 265.46 kbps, where the anchor codec cannot operate due to the maximum quantization parameter of the codec. However, from this figure it can be seen that the anchor codec’s performance gets ahead of the proposed codec’s performance at higher bitrates (from 500 kbps), which is not the target bandwidth of the proposed codec. This can be explained by the fact that the down-sampling process, which is used in the proposed codec, is a lossy phenomenon and the used enlargement algorithm (Bi-cubic) cannot fully compensate this loss of information. Figure 11 shows the results for coding “Kendo” test videos using the proposed and anchor codecs. The “Kendo” multiview video dataset is recorded under multiple artificial lighting conditions similar to the “Balloons” multiview dataset. This dataset has a progressively changing background objects with very little motion and two key moving objects in the foreground. From Fig. 11, it can be noted that the proposed codec achieves significantly higher coding

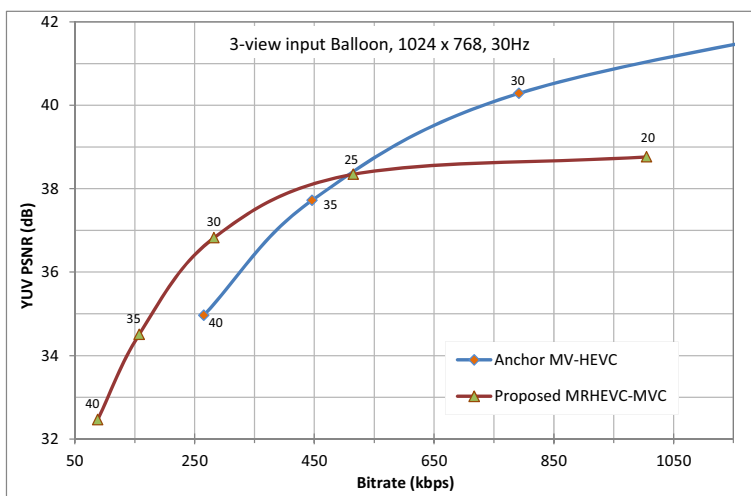


Fig. 10 Average YUV-PSNR of the anchor MV-HEVC and the proposed MRHEVC-MVC codecs for coding “Balloons” multiview videos

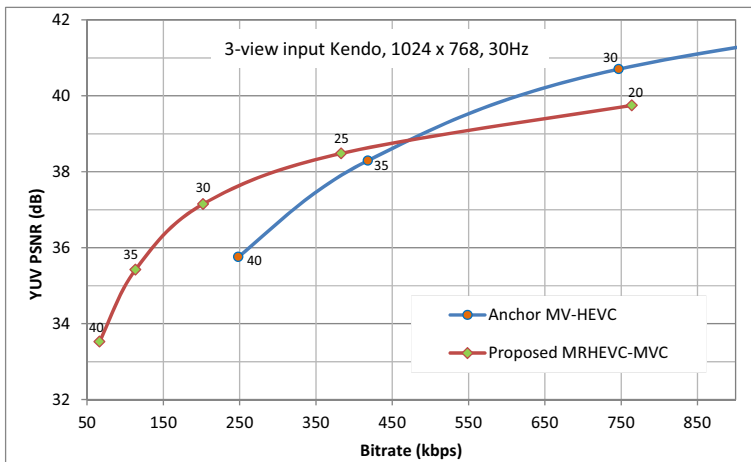


Fig. 11 Average YUV-PSNR of the anchor MV-HEVC and the proposed MRHEVC-MVC codecs for coding “Kendo” multiview videos

performance than the anchor codec at lower bitrates (1.8 to 0.2dBs higher PSNR than the anchor codec from 250 kbps to 450 kbps). The coding performance difference of the proposed codec to that of the anchor codec increases as the QP increases and it goes beyond the bitrates that the anchor codec can operate. Similar to the “Balloons” dataset’s results, the proposed codec’s coding performance for the “Kendo” dataset falls behind that of the anchor codec from 450 kbps as the bitrate increases, for the same reasons.

Figure 12 shows the results for coding the “Poznan_Street” multiview videos. The “Poznan_Street” dataset is outdoor multiview videos, captured under natural lighting condition. It contains multiple moving objects with a stationary background and fixed camera position.

From Fig. 12, it can be noted that the proposed codec outperforms the anchor codec at bitrates lower than about 1000 kbps. Its coding gain to that of the anchor codec increases as the bitrate decreases (from 0.42 dB to 1.4 dB for encoding at 800 kbps to 327.24 kbps). An

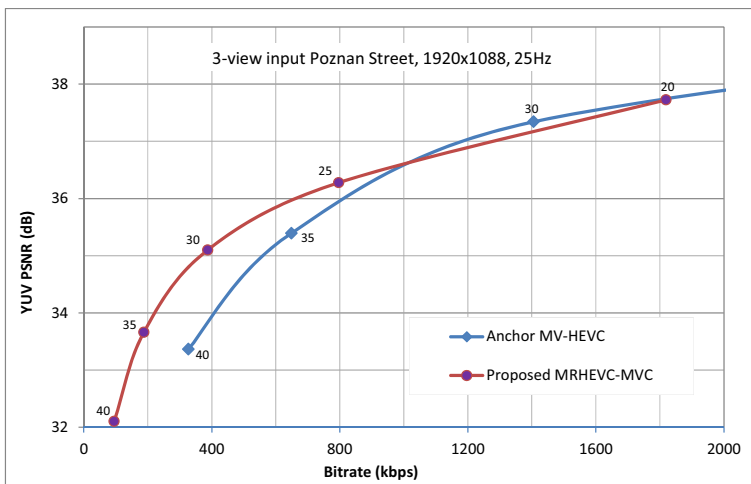


Fig. 12 Average YUV-PSNR of the anchor MV-HEVC and the proposed MRHEVC-MVC codecs for coding “Poznan_Street” multiview videos

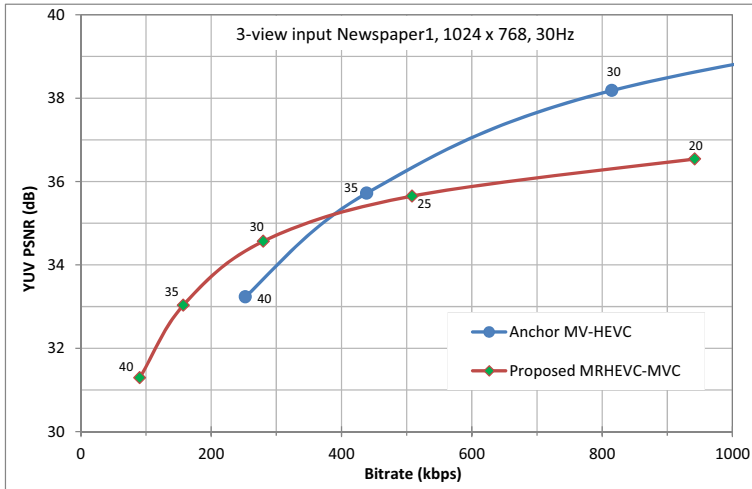


Fig. 13 Average Y-PSNR of the anchor MV-HEVC and the proposed MRHEVC-MVC codecs for coding “Newspaper1” multiview videos

average YUV-PSNR of 0.98 dB higher coding gain for the proposed codec over the anchor codec for encoding “Poznan_Street” multiview videos between 327.24 kbps and 1819.42 kbps can also be determined. Since these test videos have large still background with small moving objects, the still areas are coded by larger CTUs, which are represented by smaller number of bits. Therefore, the YUV-PSNR of the proposed codec is limited to an average of 1 dB due to the size of the small moving objects.

The “Newspaper1” multiview videos are indoor recordings with very few artificial illuminations. The scene of this video has a still background with limited movement in the foreground objects. The moving objects in these multiview videos are close to the cameras, which are simulating an interactive application such as teleconferencing. The resulting YUV-PSNR for coding the “Newspaper1” multiview videos is shown in Fig. 13. From this figure, it

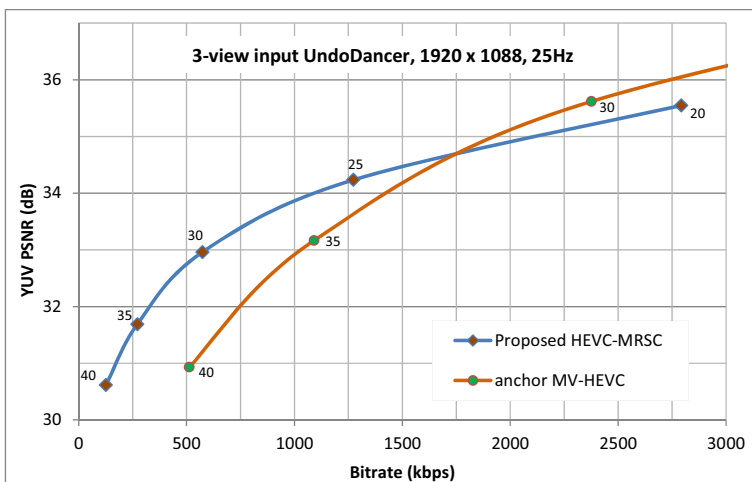


Fig. 14 Average YUV-PSNR of the anchor MV-HEVC and the proposed MRHEVC-MVC codecs for coding “Undo_Dancer” multiview videos

can be seen that the proposed codec's videos exhibit higher coding performance in terms of the average YUV-PSNR at lower bitrates and the difference in coding performance increases as the transmission bitrates decreases. The proposed MRHEVC-MVC codec gives 0.2 dB higher PSNR than the anchor codec at 370 kbps and this difference increases to 1.1 dB at 252.42 kbps. From Fig. 13, it can be seen that the anchor codec's videos exhibits higher YUV-PSNR quality than that of the proposed codec at higher bitrates, from 370 kbps, as was expected and justified earlier in the text.

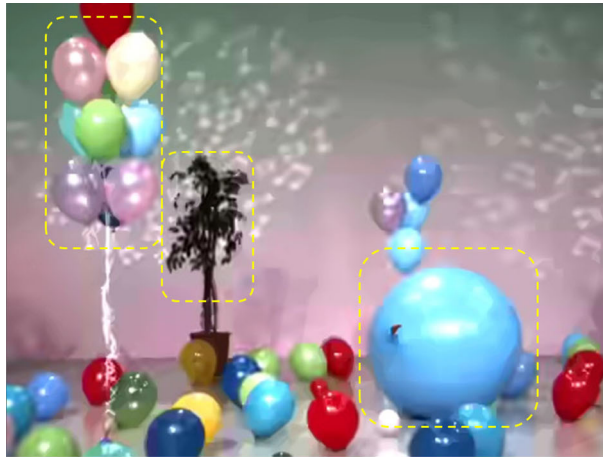
Figure 14 shows the results for coding the "Undo_Dancer" multiview video dataset. "Undo_dancer" is a computer graphic animation with scene changes representing camera motion and moving objects, where the entire scene has the same illumination.

From Fig. 14, it is evident that the proposed codec yields a higher coding performance in terms of the average YUV-PSNR than the anchor codec (almost 0.4 dB at 1500 kbps and the gap increases to 1.85 dB at around 500 kbps). Moreover, the coding performance of the anchor codec overtakes the proposed codec's performance at higher bitrates. This can be explained by the fact that the down sampling process that the proposed codec uses is a lossy process and the frame enlargement technique used by the proposed codec is unable to fully compensate this loss. Hence at higher bitrate transmission, which is not the target bitrate of the proposed codec, the anchor codec videos have higher average YUV-PSNRs. This figure represents similar trend of results for the proposed and the anchor codec, when they code "Poznan_Street", "Kendo", "Balloons" and "Newspaper1" multiview videos. Upon investigating the contents of the test videos against the coding performance of the proposed codec, it can be inferred that the proposed codec seems to yield significantly higher coding performance when videos contain moving objects or/and camera panning. The coding performance difference of the proposed codec in comparison to the anchor codec increases as the total bitrate decreases. This can be explained by the fact that the proposed mixed-resolution multiview video codec needs less frame signaling (e.g. motion vector and CTUs) to represent low resolution video frames where there is a very limited number of bits to represent the video frames.

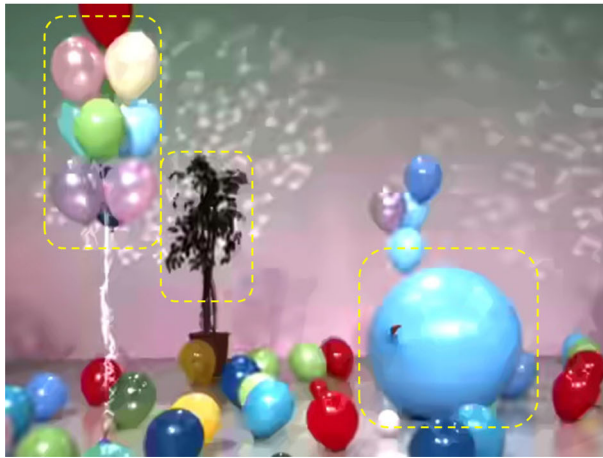
A comparison of the achieved visual quality using the proposed codec and the anchor MV-HEVC codec is depicted in Fig. 15. This figure shows a decoded middle view frame (frame number 98) of "Balloons" test videos at 304.4 kbps. From these figures, it can be seen that the proposed MRHEVC-MVC codec's image, Fig. 15a, exhibits generally higher visual quality than that of the anchor MV-HEVC's image, shown in Fig. 15b.

To enable the reader to perceive and compare the achieved visual quality using the proposed and the anchor MV-HEVC codec, snippets of the same highlighted areas of the two methods' video frames are shown in Fig. 16a, b and c. From these figures, it can be seen that the anchor codec's image exhibits significantly higher blocking artefacts around the border of the balloons, particularly the big blue balloon on the right side of the frame (as shown in Fig. 16a) and bunch of balloons on the left hand side of the image (as shown in Fig. 16b). Moreover, the plant in the background (Fig. 16c) frame of the anchor codec has lost its details and exhibits blurry, while the proposed codec has retained more details of the plant.

To further understand the coding performance of the proposed MRHEVC-MVC codec with respect to the anchor MV-HEVC codec, Bjøntegaard delta-bitrates (BD-rate) and Bjøntegaard delta-PSNR (BD-PSNR) of the decoded 'Balloons', 'Kendo', 'Poznan_Street', 'Undo_Dancer' and 'Newspaper1' multiview videos are calculated and tabulated in Tables 2 and 3. For these calculations, the piece-wise cubic interpolation method introduced in [3, 4] for five data points based interpolation polynomial, as recommended in JCTVC-B055 document [23], were used.



(a) Proposed MRHEVC-MVC



(b) Anchor MV-HEVC

Fig. 15 Decoded middle view frame number 98 from “Balloons” test videos encoded at 304.4 kb/s using **a** the proposed MRHEVC-MVC codec’s frame and **b** the anchor MV-HEVC standard codec’s frame

From Table 2, it can be seen that the proposed codec delivers superior coding performance to the standard MV-HEVC anchor codec. The average BD-PSNR Y of the proposed codec for coding “Balloons”, “Kendo”, “Poznan_Street” and “Newspaper1” multiview videos, which are camera recorded not computer graphic animated videos, is 1.201279 dB whereas, the BD-PSNR Y of the proposed codec’s “Undo_Dancer” video frames, which are computer graphic animated videos, is 2.774883 dB. Overall, the proposed codec’s videos exhibit an average BD-PSNR Y of 1.564593 dB higher with respect to the anchor codec. The proposed codec’s videos have average BD-PSNR U and BD-PSNR V of 0.090460 dB and 0.357733 dB higher with respect to that of the anchor codec’s videos, respectively. This shows a marginal BD-PSNR U and BD-PSNR V improvement, by the proposed codec, while its BD-PSNR Y is significantly higher to that of the anchor codec. Since human eyes are more sensitive to luminance than the chrominance, this implies significant improvements in terms of visual quality.

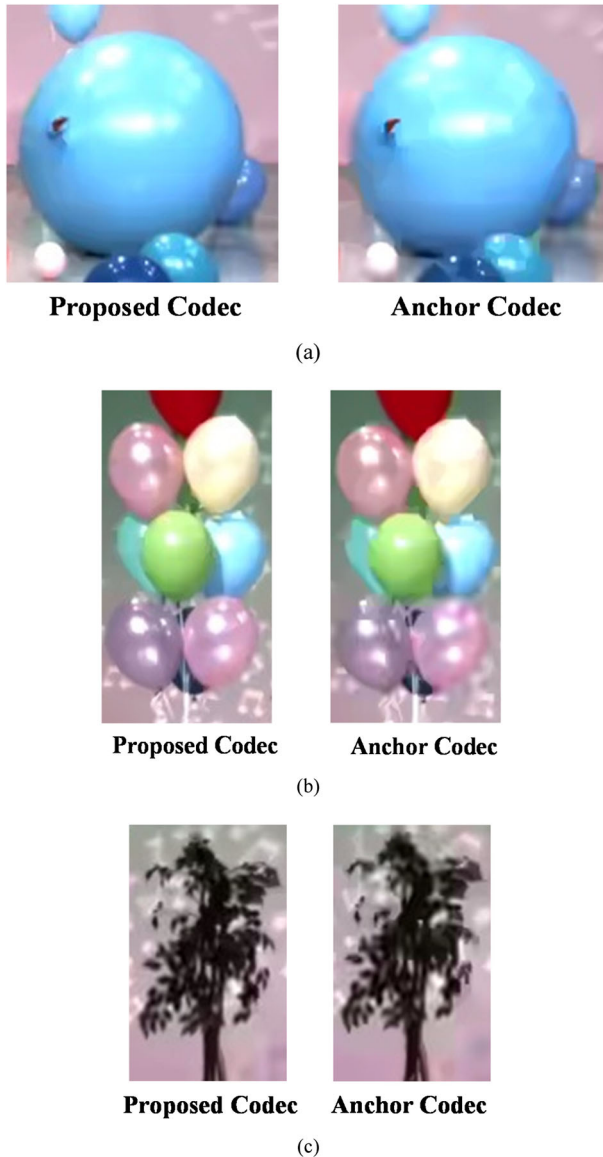


Fig. 16 Snippets of the highlighted areas of the proposed and the anchor codec's decoded frame from Fig. 15

Table 3 shows the achieved Bjøntegaard delta-bitrates (BD-rate) for the “Balloons”, “Kendo”, “Poznan_Street”, “Undo_Dancer” and “Newspaper1” multiview videos of the proposed codec with respect to that of the anchor codec. In this table, the negative values indicate that the proposed codec requires fewer bits than the anchor codec to deliver similar objective quality in terms of PSNR. From Table 3, it can be seen that the proposed codec's BD-Rate Y, BD-Rate U and BD-Rate V are on an average 38.0276 kbps, 1.86664 kbps and 8.36314 kbps lesser than that of the anchor codec's bitrates, respectively. The total average bitrate savings of the proposed codec with respect to that of the anchor codec is found to be

Table 2 BD-PSNR of the proposed codec with respect to anchor codec

Sequence	BD-PSNR (dB)		
	BD-PSNR Y	BD-PSNR U	BD-PSNR V
Balloons	1.247747	0.316506	0.109284
Kendo	1.136544	0.090357	1.537445
Poznan_Street	1.216109	0.164791	0.099425
Newspaper1	1.204719	0.056342	0.105571
Undo_Dancer	2.774883	0.005015	0.366649
Average	1.564593	0.090460	0.357733

significantly lower in terms of the BD-rate (16.0858 kbps). This implies that the proposed codec could deliver videos with similar objective qualities at significantly lower bitrates.

In order to empirically compare the coding complexity of the proposed MRHEVC-MVC codec with that of the anchor MV-HEVC, the encoding time ratio, t_E , and the encoding time percentage change, ΔT , are used. The encoding time ratio is the ratio of geometric mean of the encoding/decoding time of a multiview video sequence at a QP for the test codec to that of the anchor codec, as proposed in JCT3V-G1100 documentation [15]. The encoding time percentage change (ΔT) which is commonly used to compare the coding complexity of two codecs in terms of execution time, is calculated using Eq. 2:

$$\Delta T = \frac{t_p - t_a}{t_a} \times 100\% \quad (2)$$

where t_p and t_a are the encoding time of the proposed codec and the anchor codec, respectively, for coding a particular multiview video at a particular QP. The proposed and the anchor MV-HEVC codecs were used to encode the ‘Balloons’, ‘Kendo’, ‘Poznan_Street’, ‘Undo_Dancer’ and ‘Newspaper1’ multiview videos at different QPs. Both codecs have been run on the same Microsoft Windows 7 based personal computer, running on a 6th generation core i7 micro-processor, with 16GB of random access memory and 500GB hard-disk drive, without any dedicated graphic processing unit (no other applications, updates or background programs were running during the simulation). As suggested in JCT3V-G1100 CTC documentation, similar coding parameters: frames to be encoded, GOP size and I-frame period, except search region and reference frames, were applied to both proposed and the anchor codec when measuring

Table 3 BD-Rate of the proposed codec with respect to anchor codec

Sequence	BD-Rate (kbps)		
	BD-Rate Y	BD-Rate U	BD-Rate V
Balloons	-34.4276	4.968753	-1.548267
Kendo	-24.7975	-0.690781	-7.78663
Poznan_Street	-40.1785	-5.910673	-0.88296
Newspaper1	-27.4996	-3.438652	-9.932365
Undo_Dancer	-63.2346	-4.261847	-21.66547
Average	-38.0276	-1.86664	-8.36314

Table 4 Encoder execution time

Sequence	Encoding Time Ratio (t_E)	Encoding Time Percentage Change (ΔT)
Balloons	31.67%	-64.79%
Kendo	34.72%	-57.71%
Poznan_Street	42.51%	-51.89%
Undo_Dancer	33.34%	-62.18%
Newspaper1	37.43%	-57.87%

their execution times. The encoder time ratio (t_E) and encoding time percentage change (ΔT) of the proposed codec with respect to the anchor codec were calculated and tabulated in Table 4.

From Table 4, it can be seen that geometric mean encoding time ratio, t_E , of the proposed codec with respect to that of the anchor codec is between: 31.67 to 42.51%. This implies that the proposed codec needed much less computation cost to encode the multiview videos. This archives due to the use of the spatial resolution scaling and also the mixed resolution coding architecture by the proposed codec. From this table, it can be noticed that the encoding time percentage change, ΔT , is negative for all the encoded video sequences, which means the proposed codec saves times than the anchor codec. The encoding time percentage change is between 51.89 and 64.79%, for coding the multiview videos used in this analysis. This can be explained by the fact that the proposed HEVC-MRSVC's intermediate frames have quarter of size of the anchor codec. Hence, the proposed codecs requires less time to encode P- and B- intermediate frames than the anchor codec, which encodes full resolution P- and B- intermediate frames.

4 Conclusions

A mixed resolution HEVC based multiview video codec (MRHEVC-MVC) was presented in this paper. The proposed codec reorders the multiview video frames into a monoscopic mixed resolution videos. The proposed MRHEVC-MVC codec encodes the I-frames in full resolution and the intermediate frames in low resolution. The HEVC standard codec was modified and configured to encode frames with different resolutions in the interleaved mixed resolution multiview videos. The low resolution intermediate frames are superimposed on to full resolution zero frames. Thus, the information in the low resolution intermediate frames in the center and neighboring views are down-sampled while the frames still retain their original sizes. This reduced the complexity of the proposed codec. Experimental results show that the proposed codec outperforms the anchor HEVC codecs for low bitrate transmission at a lower computation costs. To further improve the coding performance of the proposed codec and reduce the computation costs of the proposed codec, a more efficient disparity prediction and an FPGA test-benches implementation would be beneficial.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Aflaki P, Hannuksela MM, Gabbouj M (2014) Adaptive spatial resolution selection for stereoscopic video compression with MV-HEVC: a frequency based approach. In: IEEE Int. Symp. Multimedia (ISM), pp. 267–270
2. Aksay A, Bilen C, Kurutepe E, Ozcelebi T, Akar GB, Civanlar MR, Tekalp AM (2006) Temporal and spatial scaling for stereoscopic video compression. 14th European Signal Processing Conf, Proc EUSIPCO 6(8):1–5
3. Bjontegaard G (2001) Calculation of Average PSNR Differences between RD curves. In: ITU-T SG 16, VCEG-M33, pp. 1–4
4. Bjontegaard G (2008) Improvements of the BD-PSNR model. In: ITU-T SG 16, pp. 1–2, VCEG-A111
5. Bouyagoub S, Sheikh-Akbari A, Bull D, Canagarajah N (2010) Impact of camera separation on performance of H. 264/AVC-based stereoscopic video codec. IET Electronics letters 46(5):345–346
6. Brust H, Smolic A, Mueller K, Tech G, Wiegand T (2009) Mixed resolution coding of stereoscopic video for mobile devices” in 3DTV Conf.: The True Vision-Capture, Transmission and Display of 3D Video
7. Chen Y, Wang YK, Ugur K, Hannuksela MM, Lainema J, Gabbouj M (2009) The emerging MVC standard for 3D video services. In: EURASIP Journal on Applied Signal Processing, pp. 8
8. Chung KL, Huang YH (2016) Efficient multiple-example based super-resolution for symmetric mixed resolution stereoscopic video coding. J Vis Commun Image Represent 39:65–81
9. Fehn C, Kauff P, Cho S, Kwon H, Hur N, Kim J (2007) Asymmetric coding of stereoscopic video for transmission over T-DMB,” in: Proc. 3DTV Conf., Kos Island, Greece, pp. 1–4
10. Gunatilake PD, Siegel MW, Jordan AG (1994) Compression of stereo video streams. In: Signal Processing of HDTV International Workshop on HDTV’93, pp. 173–185
11. High Efficiency Video Coding (2016) ITU-T Recommendation H.265 and ISO/IEC 23008-2
12. ITU-T and ISO/IEC JTC 1 (2010) Advanced video coding for generic audio visual service. In: Document ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)
13. Mallik B, Sheikh-Akbari A (2016) HEVC based multi-view video codec using frame interleaving technique. In: 9th Int. Conf. Developments on eSystems Engineering, DeSE, UK
14. Merkle P, Smolic A, Muller K, Wiegand T (2007) Efficient prediction structures for multiview video coding. IEEE Trans Circuits Syst Video Technol 17(11):1461–1473
15. Muller K, Vetro A (2014) Common test conditions of 3DV core experiments. In: ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V G1100, pp. 1–7
16. Perkins MG (1992) Data compression of stereopairs. IEEE Trans on Communications 40(4):684–696
17. Said H, Sheikh-Akbari A (2013) H.264/AVC Based multi-view video codec using the statistics of block matching. 55th international symposium ELMAR, pp. 97–100
18. Said H, Sheikh-Akbari A, Moniri M (2013) An adaptive reference frame re-ordering algorithm for H.264/AVC based multi-view video codec” in International conference EUSIPCO, pp. 1–5
19. Said H, Moniri M, Chibelushi CC (2017) Prediction architecture based on block matching statistics for mixed spatial-resolution multi-view video coding. EURASIP Journal on Image and Video Processing 2017:15
20. Sansli DB, Ugur K, Hannuksela MM, Gabbouj M (2014) Inter-view motion vector prediction in multiview HEVC. In: 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), pp. 1–4
21. Saygili G, Gürlür CG (2011) Evaluation of asymmetric stereo video coding and rate scaling for adaptive 3D video streaming. IEEE Trans on Broadcasting 57(2):593–601
22. Schwarz H, Schierl T, Marpe D (2013) Block structure and parallelism. In: High efficiency video coding (HEVC): Algorithms and Architectures, Integrated Circuit and Systems, USA: Springer, ch. 3, sec. 3.2.5, pp. 49–90
23. Senzaki K (2010) BD-PSNR/Rate computation tool for five data points. In: ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCTVC-B055, pp. 1–3
24. Sheikh-Akbari A, Bagheri Zadeh P (2015) Wavelet Based Image Enlargement Technique. In: Int. Conf. Global Security, Safety, and Sustainability. Springer International Publishing, pp. 182–188
25. Sheikh-Akbari A, Said H, Moniri M (2012) Effect of inter-camera angles on the performance of an H. 264/AVC based multi-view video codec. In: PCS, pp. 109–112
26. Smolic A (2011) 3D video and freeviewpoint video: From capture to display. Pattern Recogn 44(9):1958–1968
27. Stelmach LB, Tam WJ, Meegan D, Vincent A (2000) Stereo image quality: effects of mixed spatio-temporal resolution. IEEE Trans Circuits and Systems for Video Technology 10(2):188–193
28. Sullivan GJ, Boyce JM, Chen Y, Ohm JR, Segall CA, Vetro A (2013) Standardized extensions of high efficiency video coding (HEVC). IEEE Journal of Selected Topics in Signal Processing 7:1001–1016
29. Vetro A, Wiegand T, Sullivan GJ (2011) Overview of the stereo and multiview video coding extensions of the H. 264/MPEG-4 AVC standard. Proc of the IEEE 99(4):626–642
30. Visual Networking Index Cisco (2016) The zettabyte era: trends and analysis. Cisco white paper

31. Zhang Y, Jiang GY, Mei Y, Yo SH (2006) An Approach to Multi-modal Multi-view Video Coding. Proc of International Conf Signal Processing (ICSP) 2:1405–1408
32. Zhang Y, Jiang GY, Yu M, Ho YS (2009) Adaptive multiview video coding scheme based on spatiotemporal correlation analyses. ETRI J 31(2):151–161



Bruhanth Mallik has a MSc degree in Electronic Engineering from University of Portsmouth and currently he is a PhD candidate at Leeds Beckett University. His research is in the direction of studying and developing efficient multiview video coding techniques for bandwidth limited transmission, within HEVC's framework.



Akbar Sheikh-Akbari is a Senior Lecturer in School of Computing, Creative Technologies & Engineering at Leeds Beckett University. He has a BSc (Hons), MSc (distinction) and PhD in Electronic and Electrical Engineering. After completing his PhD at Strathclyde University, he joined Bristol University to work on an EPSRC project in stereo/multi-view video processing. He continued his career in industry, working on real-time embedded video analytics systems. His main research interests include: signal processing, hyperspectral image processing, source camera identification, image/video forgery, image hashing, biometric identification techniques, assisted living technologies, compressive sensing, camera tracking using retro-reflective materials, standard and non-standard Image/video codecs, e.g. H.264 and HEVC, multi-view image/video processing, colour constancy techniques, resolution enhancement methods, edge detection in low SNR environments and medical image processing.



Ah-Lian Kor specialises in software development, web applications, data analytics, and AI. She is active in sustainable IT, intelligent systems, decision support systems, and data center research. She forges an industrial collaboration with Drax Power Station which aims to develop and evaluate a virtual environment (in the form of an expert system) to support decision making in fault diagnosis for the centrifuge system for Drax Power Station. She is an Intelligent System Academic Expert for the Knowledge Transfer Partnership Project with Premier Farnell (a global electronics engineering company). Ah Lian is a Project Team Committee Member for several EU projects (e.g. ALIOT, GreenCo, PERCCOM, etc.). She has chaired sessions in reputable international conferences (e.g. FTC 2016), a publicity chair for CyberSciTech 2017, sits in Technical Programme Committee for journals and conferences (e.g. ICAART, SEEDS, IARIA Intelligent Systems, IEEE Computing Conference, etc.).